Most materials in this course are based on materials from http://aima.cs.berkeley.edu/ and http://ai.berkeley.edu


# 0. Course Details & Introduction

COMP 3270
Artificial Intelligence

Dirk Schnieders

# 0. Course Details & Introduction - Outline

- Course Details
  - Prerequisite
  - You will not learn …
  - Staff
  - Syllabus
    - Topics
    - Schedule
  - Assessment
  - Plagiarism
  - Slides
  - Zoom
  - Discussion Forum
  - Textbook
- Introduction
  - Intelligence vs AI
  - Goal of AI
  - Rational Agent
  - Machine Learning
  - History of AI
    - Turing Award Winners
    - Milestones
  - The State of the Art
  - Risks and Benefits of AI
    - Superhuman AI
    - The Gorilla Problem

# Prerequisite

- COMP2119 Introduction to Data Structures and Algorithms (or equivalent)
  - Mathematics foundation
    - Growth of functions, recurrences, time and space complexity
  - Basic data structures and algorithms
    - Array, linked list, stack, queue, trees, hashing, graphs, sorting and searching algorithms
  - Advanced data structures and algorithms
    - Balanced tree, heap, priority queue, sorting in linear time

# You will not learn ...

- Machine Learning: Perceptron, SVM, Linear Regression, ...
  - Consider taking: comp3314 Machine Learning
- Deep Learning (CNN, RNN, GAN, Deep RL)
  - Consider taking: comp3340 Applied deep learning, and/or
  - Consider taking: comp3317 Computer vision
- To work on bigger applied projects
  - Consider taking: comp3414 Experiential learning on AI and robotics
- R language with a focus on statistical modeling
  - Consider taking: comp3354 Statistical learning
- Advanced NLP models: Take COMP3361

# Staff

- Instructor: Dirk Schnieders
  - Email: sdirk@cs.hku.hk
  - Consultation hours: Tue 2:30 - 3:30 & by appointment
- TAs
  - Cao Yukang
    - Q1, A1, A2, Forum admin during time: [start of course, Q1]
    - Email: ykcao@cs.hku.hk
    - Consultation hours
      - Thursday 2:30-3:30, Room 411 Chow Yei Ching Building
  - Guo Qiushan
    - Q2, A3, A4, Forum admin during time: [Q1, final exam]
    - Email: u3007956@connect.hku.hk
    - Consultation hours: Thu 2:30-3:30, Room P501 - Graduate Hall

# Syllabus - Topics*

*subject to change

- Part I: Search and Planning
  - Search (Uninformed, Informed, Local, CSPs, Adversarial)
  - MDPs (Value Iteration, Policy Iteration)
  - RL (Temporal Difference Learning, Q Learning)
- Part II: Probabilistic Reasoning
  - MMs (Probability Review, Markov Chain, Mini-Forward Algorithm)
  - HMMs (Forward Algorithm, Particle Filtering)
  - BNs (Inference, Sampling)
- Part III: Intro to Machine Learning
  - NLP or another ML application (send me an email with topic requests)

# Syllabus - Schedule*

*subject to change



| Week | Tue (1h) | Topic* | Fri (2h) | Topic* |
|---|---|---|---|---|
| 1 (1 Sep - 2 Sep) | | | 2 Sep | Course Details & Introduction |
| 2 (5 Sep - 9 Sep) | 6 Sep | Uninformed Search | 9 Sep | Informed Search |
| 3 (12 Sep - 16 Sep) | 13 Sep | Local Search | 16 Sep | CSPs |
| 4 (19 Sep - 23 Sep) | 20 Sep | Adversarial Search | 23 Sep | Adversarial Search |
| 5 (26 Sep - 30 Sep) | 27 Sep | Markov Decision Processes | 30 Sep | Markov Decision Processes |
| 6 (3 Oct - 7 Oct) | 4 Oct | no class | 7 Oct | Reinforcement Learning |
| 7 (10 Oct - 14 Oct) | 11 Oct | no class | 14 Oct | no class |
| 8 (17 Oct - 21 Oct) | 18 Oct | Quiz 1 | 21 Oct | Reinforcement Learning |
| 9 (24 Oct - 28 Oct) | 25 Oct | Markov Models | 28 Oct | Markov Models |
| 10 (31 Oct - 4 Nov) | 1 Nov | Hidden Markov Models | 4 Nov | Hidden Markov Models |
| 11 (7 Nov - 11 Nov) | 8 Nov | Hidden Markov Models | 11 Nov | Bayes Nets |
| 12 (14 Nov - 18 Nov) | 15 Nov | Bayes Nets | 18 Nov | Bayes Nets |
| 13 (21 Nov - 25 Nov) | 22 Nov | NLP | 25 Nov | NLP |
| 14 (28 Nov - 30 Nov) | 29 Nov | Quiz 2 | | |

# Assessment - Weight

- Exam: 50%
- Quiz 1: 8%
- Quiz 2: 8%
- Assignment 1: 10%
- Assignment 2: 8%
- Assignment 3: 8%
- Assignment 4: 8%

mark = Q1 + Q2 + A1 + A2 + A3 + A4 + exam
grade = m(mark)
the mapping function m() will not be published

# Assessment - Quiz 1 & 2

- Written, closed book, calculator allowed
- Plagiarism cases will receive 0 marks and referred to program director
- 0 marks unless sick leave certificate (on day of test) or official leave from university
- Quiz 1: Scope: Everything covered in course
- Quiz 2: Scope: Everything covered in course since Quiz 1

# Assessment - Assignments

- 4 Programming Assignments
  - Python 3, Suggestion: Install Anaconda 3
  - ~2 weeks time, no deadline extension
    - Late submission, after the deadline, will result in 0 marks
    - Submission of wrong files will result in 0 marks
      - Please check very carefully
  - Release - Deadline (subject to change)
    - A1: 13 Sep - 27 Sep
    - A2: 27 Sep - 11 Oct
    - A3: 25 Oct - 8 Nov
    - A4: 11 Nov - 25 Nov

# Assessment - Final Exam

- 2 hours
- Details to be announced

# Assessment - Results

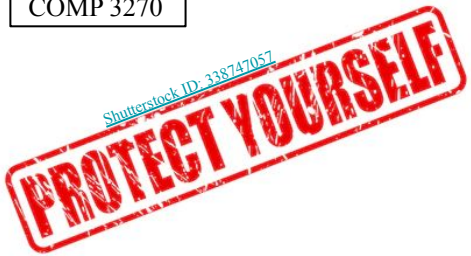
- You must check for the correctness of your in-class assessment on Moodle
  - If you find any problem, send an email to our TA immediately

# Plagiarism

- What is Plagiarism ?
  - https://tl.hku.hk/plagiarism/
- First attempt
  - Written warning
  - Zero marks
  - Referred to program director
- Second attempt
  - Referred to University Disciplinary Committee
  - Published reprimand
  - Suspension of study
  - Expulsion

# Plagiarism

- We will treat your submission as plagiarism if we have sufficient evidence
  - Source will be reported and will also receive 0 marks
- You will receive a formal warning by the department in writing




# Prevent Plagiarism

- Never copy/paste anything **without citing**
  - If you copy existing work you must clearly state and cite appropriately
  - Solutions to assignments may be found on the internet
    - To prevent plagiarism, don't search for them
- Never look at existing solutions
  - Work on the problem yourself
- Never ask your friend to see his/her solution
  - Note that the source will also be punished with 0 marks

# Slides

- Available on Moodle on the day before the lecture
- Some slides include comments/solutions that I use during the lecture
- Use the provided materials responsibly
  - Don’t share with others
  - Do not upload to internet
    - I do not hold copyright for most materials

# Recording

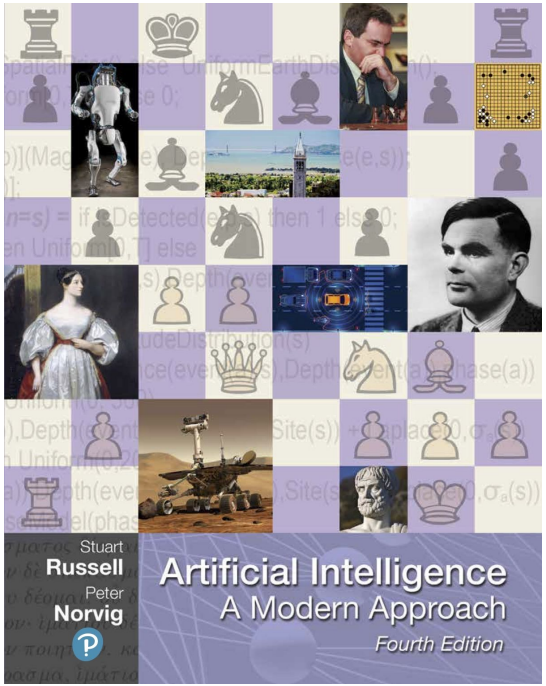
- All lectures will be held face2face only
- Recordings might be uploaded
  - Do not rely on the recording, there could be a technical issue or I could forget to switch it on
  - Please remind me to switch on the recording

# Discussion Forum

- If you have questions about course materials, please **use the forum**
  - We will try to answer within 48 hours
  - Please help to answer questions by others
- If you contact us by email please use your university email and include your university number in the email

# Textbook

- Artificial Intelligence: A Modern Approach (4th Edition)
  - http://aima.cs.berkeley.edu/



# Introduction

# Intelligence

- We call ourselves Homo sapiens
  - Latin: Wise Man
  - Intelligence is important to us
- For thousands of years, we have tried to understand how we humans think
- Intelligence is most widely studied in humans, but has also been observed in animals and in plants

# Artificial Intelligence

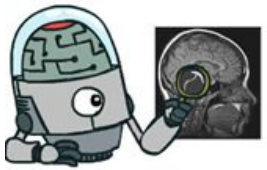
- AI goes further than just understanding intelligence
  - It attempts to build intelligent entities
- Is AI science, or is it engineering?
  - AI's science goal
    - To understand the principles and mechanism that account for intelligent action
  - AI's engineering goal
    - To design intelligent artifacts that can survive and operate in the physical world and solve problems of considerable scientific difficulty at high levels of competence

# Goal of AI

- Should we create machines that …

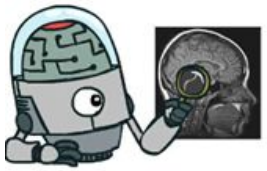
A. Think Humanly
B. Act Humanly
C. Think Rationally
D. Act Rationally

# A. Think Humanly

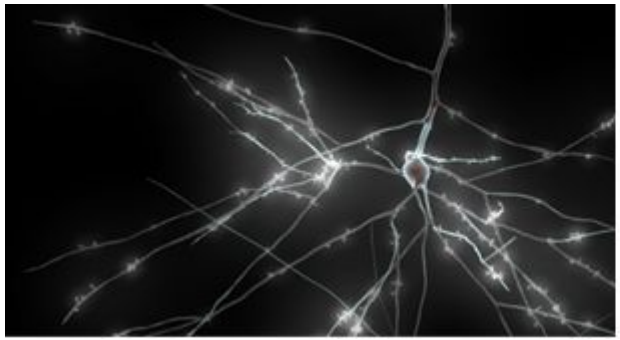


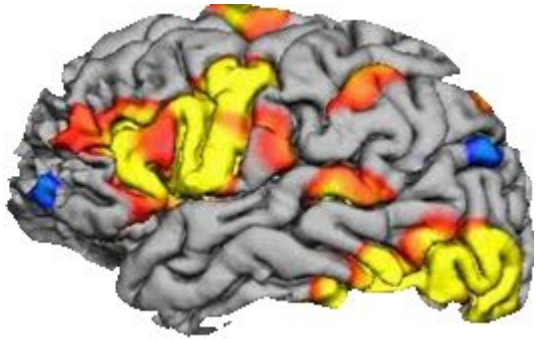
- Researchers find out how we think by
  - Introspection
  - Psychological experiments
  - Brain imaging
- Cognitive science constructs precise and testable theories of the human mind
  - E.g., express a theory as a computer program and compare input-output behaviors to a human
  - If there is a match, some of the programs mechanism could also be operating in humans

# A. Think Humanly



- The human brain is one of the great mysteries of science
  - How does our brain process information?
- The brain consists of nerve cells (aka neurons) and the collection of these simple cells leads to thought, action and consciousness
- The recent development of functional magnetic resonance imaging (fMRI) provides neuroscientists with details of brain activities





# A. Think Humanly

- Brains and digital computers have somewhat different properties
- A crude comparison of the raw computational resources

| | Supercomputer | Personal Computer | Human Brain |
|---|---|---|---|
| Computational units | $10^6$ GPUs + CPUs | 8 CPU cores | $10^6$ columns |
| | $10^{15}$ transistors | $10^{10}$ transistors | $10^{11}$ neurons |
| Storage units | $10^{16}$ bytes RAM | $10^{10}$ bytes RAM | $10^{11}$ neurons |
| | $10^{17}$ bytes disk | $10^{12}$ bytes disk | $10^{14}$ synapses |
| Cycle time | $10^{-9}$ sec | $10^{-9}$ sec | $10^{-3}$ sec |
| Operations/sec | $10^{18}$ | $10^{10}$ | $10^{17}$ |

- Would we be able to achieve the brain's level of intelligence with a computer of unlimited capacity?

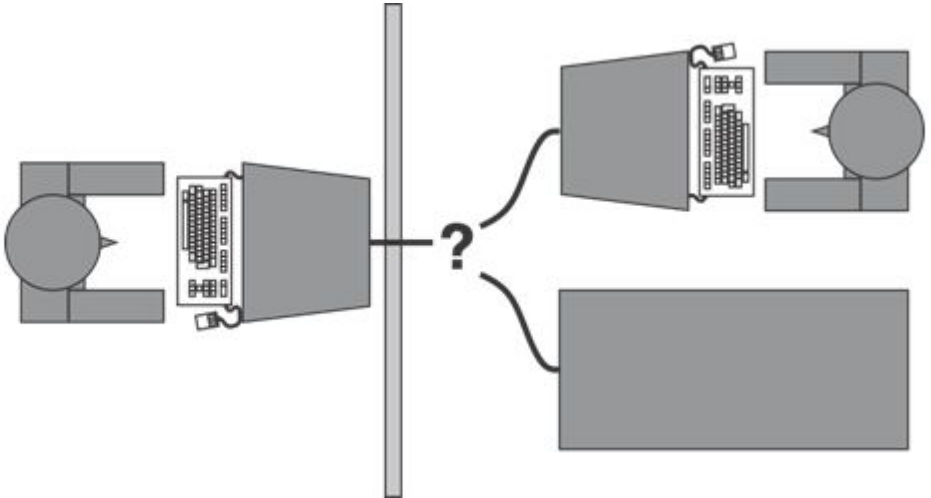
# Should we create machines that ...

A. Think Humanly
B. Act Humanly
C. Think Rationally
D. Act Rationally

# B. Act Humanly

- The Turing Test (aka Imitation Game) was designed to provide a definition of intelligence
- A computer passes the test if a human interrogator, after posing some questions, cannot tell whether the response come from a human or a computer

Human interrogator

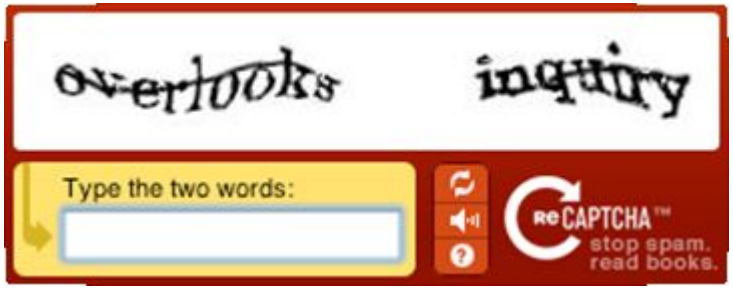


# B. Act Humanly

- The underlying principles of intelligence are more important than to duplicate an exemplar
- Consider another field: Artificial Flight
  - The Wright brothers succeeded because they stopped imitating birds and started using wind tunnels and learn about aerodynamics
  - It was not their goal to make “machines that fly so exactly like pigeons that they can fool even other pigeons”

# B. Act Humanly

- A reverse Turing test is a Turing test in which the objective or roles between computers and humans have been reversed. Interrogator is a computer. Interrogatee is a human.
  - Example: CAPTCHA
    - It is program that can generate and grade tests that
      - most humans can pass, but
      - current computer programs cannot pass



# C. Think Rationally

- What are the laws that guide and underlie our thinking?
- Greek schools developed various forms of logic
  - Notation and rules of derivation for thoughts
  - Example: Socrates is a man; all men are mortal; therefore, Socrates is mortal
- By 1965, programs existed that could (in principle) solve any solvable problem described in logic notation
- Problems with this approach
  - How to take informal knowledge and state it in formal terms? How about uncertainty?

# D. Act Rationally

- Act so as to achieve the best outcome or, when there is uncertainty, the best expected outcome
- Advantages over the other approaches
  - More general than the laws of thought approach because correct inference is just one of several possible mechanisms for achieving rationality
  - The standard of rationality is mathematically well defined. Human behavior, on the other hand, is well adapted only for one specific environment
- In this course we will focus on the general principles of rational agents and how to build them
  - An agent is something that perceives and acts

# Rational Agent

- This course is about designing rational agents
- For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance
  - Computational limitations make perfect rationality unachievable
  - Design best program for given machine resources

# Machine Learning

- An agent is learning if it improves its performance on future tasks
- Why would we want an agent to learn? If the design of an agent can be improved, why not design the agent with that improvement to begin with?
  - Cannot anticipate all possible situations
  - Cannot anticipate all changes over time
  - Don’t know how to program some solutions

# History of AI

# History of AI - Turing Award Winners

- Marvin Minsky (1969)
- John McCarthy (1971)
- Edward Feigenbaum and Raj Reddy (1994)
- Judea Pearl (2011)
- Yoshua Bengio, Geoffrey Hinton, and Yann LeCun (2019)

# History of AI - Milestones

- Inception (1943 - 1956)
- Early Enthusiasm (1952 - 1969)
- A dose of reality (1966 - 1973)
- Expert systems (1969 - 1986)
- Return of NN (1986 - present)
- Probabilistic reasoning (1987 - present)
- Big data (2001 - present)
- Deep Learning (2011 - present)

# The State of the Art

- Publications
  - AI papers increased 20 fold between 2010 to 2019 to 20,000 a year
- Conferences
  - Attendance of NeurIPS increased 800% since 2012 to 13,500
- Industry
  - AI startups in the US increased 20 fold to over 800 from 2010 to 2019
- Internationalization
  - China publishes more AI papers per year then US and about as many as Europe
  - In citation weighted impact, US is ahead by 50% vs. China

# The State of the Art

- Vision
  - Error rates for object detection improved from 28% to less than 2%
- Speed
  - Training time for image recognition dropped by a factor of 100 in last 2 years
  - Amount of computing power used in top AI applications is doubling every few month
- Humans vs. AI
  - AI is better in chess, go, pojer, pac-man, jeopardy!, object detection, speech recognition in limited domain, chinese-to-english in restricted domain, Quake III, Dota 2, StarCraft II, many Atari games, Skin cancer detection, prostate cancer detection, protein folding, ...

# Benefits of AI

First solve AI, then use AI to solve everything else.

Demis Hassabis, Google DeepMind

# Risks of AI

- Lethal autonomous weapons
- Surveillance
- Biased decision making
- Impact on employment
- Safety-critical applications
- Cybersecurity

# Risks of AI - Superhuman AI

- Most experts agree that we will eventually be able to create a superhuman AI
  - An intelligence that far surpases human ability

# Risks of AI - The Gorilla Problem

- About seven million years ago, a now-extinct primate evolved
  - one branch led to gorillas
  - another to humans
- Today the gorillas are probably not too happy about the human branch
  - They have no control over their future

# Risks of AI - The Gorilla Problem

- If the gorilla problem is the result of developing AI then we should stop working on it
- If superhuman AI were a black box from outer space, we should be careful in opening the box
  - But it is not, **we** design the AI systems
  - If AI does end up taking control, it would be a design failure
- We need to understand the source of potential failure
  - Philosophical foundations of AI
  - Maybe the most important area of AI research

# Reference

- Textbook: Chapter 1